4

Critical interactions in the globular bottom of influenza hemagglutinin

Abstract

Influenza hemagglutinin (HA) mediates fusion of the endosomal and viral membranes during viral invasion. Single-particle experiments have indicated that half of the available HAs are non-productive in the fusion process, presumably by a collapse of the fusion-active subunit (HA2) before insertion of its amphipathic fusion peptides into the target membrane. As such, the balance between the rates of insertion and collapse seems to determine HA productivity. By mechanical unfolding of the globular bottom of HA2, we investigated which molecular interactions in HA control the rate of collapse. We find Arg163 as a critical residue within a network of stabilizing salt bridges in the core of the globular bottom. Preliminary results from single-particle experiments are consistent with the hypothesis that the number of non-productive HAs increases when mutations disturb this network. However, further experiments are needed to examine if other effects of the mutations might impact fusion efficiency. The regulation of fusion efficiency through targeting conserved residues, such as the ones found here, can potentially be used as a generic antiviral strategy.

Boonstra S, Blijleven JS, Ivanovic T, Onck PR, van Oijen AM, van der Giessen E. Critical interactions in the globular bottom of influenza hemagglutinin. *In preparation*.

4.1 Introduction

The traffic of biological cargo between cells and between intracellular compartments relies on the presence of pores, which can be either inherent to the cell membrane or formed by membrane fusion. This latter mechanism unifies individual compartments by joining together their delineating membranes. Although lipid bilayer fusion is thermodynamically favourable, the end states are separated by a high kinetic barrier.[11] Various biological processes have evolved their own specialized fusion proteins to overcome this barrier within a biologically relevant timescale. Examples include synaptic vesicle fusion in neuron transmission by SNARE proteins,[39] fusion in the endoplasmic reticulum by GTPases[22] and fusion by viral fusion proteins during viral entry.[19,27] Of the latter, influenza hemagglutinin (HA) is the most thoroughly characterized. Both its prefusion[47] and postfusion[8] structure were the first to be crystallized of all fusion proteins. The amount of knowledge already acquired about HA, and the strong resemblance of its postfusion structure with those of other fusion proteins, has made it a popular model system to study the operating mechanism of other fusion proteins.[33,43] A hypothesized pathway for HA-mediated fusion has been inferred from the conformational differences between the pre- and postfusion crystal structures, backed up by biochemical experiments.[5] HA is a homotrimeric glycoprotein that is embedded in the membrane of a virus particle through a transmembrane domain. Upon cleavage, each monomer reorganizes into two domains, HA1 and HA2, that are connected by disulfide bonds, as shown in Figure 4.1a. The globular HA1 domain can bind to receptors on the plasma membrane (Figure 4.1b), after which the virus particle is internalized by endocytosis. Acidification of the endosome to a pH below 6 activates the fusion process, by dissociating HA1 and exposing hydrophobic fusion peptides at the N-terminal of the fusion domain, HA2 (Figure 4.1c).[14] The pathway shown in the top row of Figure 4.1 is characterized by a timely transition of the previously unstructured B-loop of HA2 into a helical coiled coil, extending the central coiled coil. This facilitates insertion of the fusion peptide into the target membrane (Figure 4.1d).[9,44] The putative 'extended intermediate' that is formed this way, bridges the two membranes. Subsequent collapse of the globular bottom of HA2 transforms the protein into a hairpin that is hypothesized to pull the membranes together in a zipper-like fashion (Figure 4.1e) for productive membrane fusion (Figure 4.1f).[10,37]

A combination of single-particle fusion kinetics experiments[36] and stochastic modeling of HA activation and insertion has led to the notion that a large fraction of HAs does not participate productively in the fusion process.[25] Instead, fusion is mediated by only a certain number of productively inserting HAs that can jointly provide enough energy to make the membranes fuse.[24] The number of HAs required for fusion, their activation rate and the fraction of non-productive HAs varies between virus strains. Two HA strains that show a marked difference in these properties are H1 and H3,[36] which belong to two distinct phylogenetic groups of influenza viruses. This variety in fusion properties between

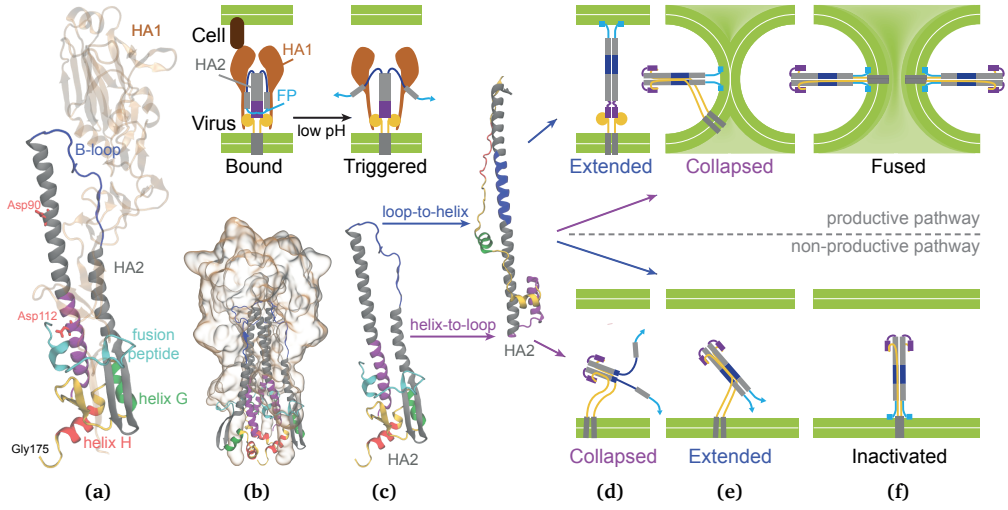


Figure 4.1: *a) Structure of one monomer of hemagglutinin (HA) in cartoon representation, with HA1 in light brown and colour-coded regions in HA2: 1-23 cyan (fusion peptide); 24-57 gray; 58-75 blue (B-loop); 76-105 gray; 106-128 purple (hinge region); 129-175 yellow (globular bottom), with 146-154 green (helix G) and 163-171 red (helix H). Residues Asp90 and Asp112 are indicated in licorice representation in red. b-f) Proposed productive and non-productive fusion pathways. b) HA binds to the target membrane on sialic-acid receptors (dark brown). c) Low pH induces dissociation of HA1 subunits (light brown) and release of the fusion peptide (cyan). In the proposed productive pathway (top row, d), the loop-to-helix transition in the B-loop (blue) enables the fusion peptides to insert into the target membrane before (e) the collapse of the trimer in the 'hinge' region (purple), unfolding and zippering up of the globular bottom (yellow), leading to (f) fusion. The hypothesized non-productive pathway (bottom row, d) is induced by a premature collapse of the hinge region (purple) and unfolding and zippering up of the globular bottom before insertion of the fusion peptides in the target membrane. Instead, after extension of the central coiled coil (e), the fusion peptides are inserted into the viral membrane (f), effectively inactivating the protein.*

strains has been proposed to be a host-specific evolutionary mechanism to optimize the location of the fusion event as close as possible to the cell nucleus, at the risk of detection by the host immune system and lysosomal degradation.[12,25,41] However, the molecular mechanism that the virus employs to tune HA productivity remains unclear.

We here explore a possible role for the globular bottom of HA2 in governing fusion productivity, as illustrated by the non-productive pathway of HA rearrangements in the bottom row of Figure 4.1. The idea is that before the fusion peptides can insert into the target membrane, the globular bottom collapses (Figure 4.1d) and the zippering mechanism starts (Figure 4.1e). Such a premature collapse effectively inactivates the protein by insertion of the fusion peptides into the viral membrane (Figure 4.1f).[45] The hypoth-

esis that the productivity of individual HAs is determined by the stability of the globular bottom is consistent with computer simulations of the pre- to postfusion transition of HA using a structure-based model, where it was found that all other HA2 rearrangements ensue rapidly as soon as the globular bottom has unfolded.[30]

To test this hypothesis, we investigated the mechanical stability of the globular bottom of two virus strains using steered molecular dynamics simulations. We find that a network of salt bridges in the core of the globular bottom provides stability against the applied load in mechanical unfolding simulations. Breaking of a particular salt bridge within this network is a critical event during the unfolding. With this knowledge, mutant viruses are designed that do not form the critical salt bridge and which unfold significantly faster than the wild-type in the simulations. Preliminary results from single-particle fusion kinetics experiments (see Appendix 4.A) show decreased fusion efficiency for these destabilizing mutations.

4.2 Methods

4.2.1 Simulation setup

Steered molecular dynamics simulations were used to mechanically unfold the globular bottom of HA from two influenza strains, H1 (A/Puerto Rico/8/34 (PR8) H1N1) and H3 (A/Aichi/68 (X31) H3N2, PDB code 1HGF[40]). These strains were selected because there is data on the fraction of productive HAs for each of them.[25,36] The structure of this specific strain of H3 is known for HA2 residues 1-175, which can be used as a starting point for the simulations. However, the crystal structures in the PDB database for H1 of this strain contain at most the residues 1-160 of HA2 (PDB code 1RU7). To obtain a structure of HA2 up to residue 175 for the simulations, a homology model of the H1 sequence of HA2 onto the structure of H3 (PDB code 1HGF) was made using the SWISS-Model web server.[4] A comparison of the modeled structure with the incomplete structure of H1 is shown in Figure 4.B.1 in Appendix 4.B.

All simulations were carried out in GROMACS 4.6,[21] using the CHARMM36 force field[3] and the standard TIP3P water model.[26] As our previous work has shown,[7] this water model yields more realistic propensities for folded structures in combination with the CHARMM36 force field than the default CHARMM TIP3P water model.[32] Langevin integration was used with a friction coefficient of $1\,\mathrm{ps}^{-1}$ to maintain the temperature at 300 K. In the case of NPT simulations, the atmospheric pressure was maintained using Berendsen pressure coupling with a coupling coefficient of $1\,\mathrm{ps}^{-1}$. The integration time step was 2 fs, while all bonds including hydrogen atoms were constrained using P-LINCS[20] while keeping the waters rigid using SETTLE.[34] Long-range electrostatics was treated with the particle-mesh Ewald (PME) approximation, using a grid spacing of 1.2 Å and a direct space cut-off of 9 Å. The van der Waals interactions were cut-off at 14 Å, in accordance

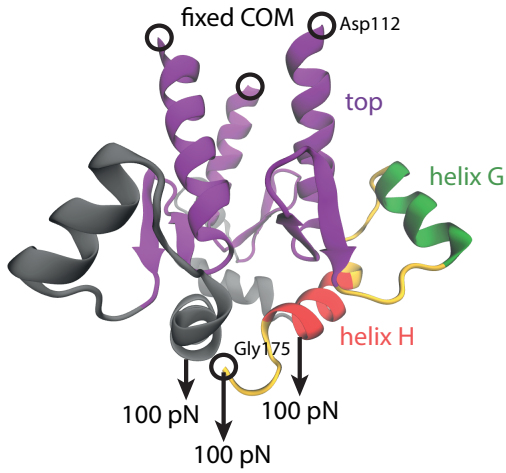


Figure 4.2: *Starting configuration and pulling groups for globular bottom unfolding of* $HA2_{112}$*. The center of mass (COM) of each of the Asp112 residues at the N-terminus are fixed at their starting coordinates. A pulling force of* 100 pN *is applied to each of the C-terminal residues Gly175 individually, in the direction parallel to the trimeric three-fold symmetry axis. The 'top' part (purple), Helix G (green) and H (red) are indicated.*

with our previous work.[7]

The starting configuration and pulling conditions for the mechanical unfolding of the globular bottom are shown in Figure 4.2. To reduce computing time, the simulations were carried out on a truncated model of prefusion HA2 ($HA2_{112}$, containing residues 112 to 175; see Figure 4.1a). The centers of mass of the N-terminal residues Asp112 were constrained to their initial positions, while the C-terminal residue Gly175 was pulled for each chain independently with a constant force. The direction of the forces is parallel to the axis of three-fold symmetry of the HA trimer.

Due to the truncation to $HA2_{112}$, the fusion peptides were removed from their positions between the helices near Asp112 in the central coiled coil (marked red in Figure 4.1a). Their absence could potentially introduce instabilities in the coiled coil during the pulling simulations, which would erroneously influence the behaviour of the globular bottom. So, in order to find the new equilibrium positions of these helices in the absence of the fusion peptides, an equilibrium MD simulation was carried out on a slightly larger model of HA2 which includes more of the coiled coil but not the fusion peptides, ($HA2_{90}$, containing residues 90 to 175; see Figure 4.1a). This showed that the average distance between the centers of mass of Asp112 residues decreased quickly, from the initial separation of about 1.5 nm in the crystal structure, to an average equilibrium distance of approximately 1.2 nm in the simulation without the fusion peptides. Accordingly, configurations with the centers

of mass of all Asp112 residues separated between 1.15 nm and 1.25 nm were taken from the simulation and were truncated at residue Asp112 to be used as initial configurations for simulations on $HA2_{112}$. An example of such a configuration is shown in Figure 4.2. Initial configurations for pulling simulations with mutations in $HA2_{112}$ were obtained using the same procedure, starting from $HA2_{90}$ models that included the respective mutations. Mutations were applied to the protein structures using PyMol.[42]

Another effect of the truncation to $HA2_{112}$ is that the helices that belong to the central coiled coil are now considerably shorter, thus decreasing their stability with respect to full-length HA2. To preclude unwanted effects from coiled-coil unfolding, constraints were applied that reinforce the helical hydrogen bonds. These constraints were implemented by adding harmonic potentials between each backbone nitrogen atom and the carboxyl carbon atom three residues further in the helix, with a force constant of 1000 kJ/(mol nm) and with the distance between the atoms in the starting configuration as the reference distance.

The applied pulling force should be high enough to unfold the globular bottom within a reasonable amount of computing time, while at the same time, it should be low enough to give stochastic processes in the globular bottom time for it to unfold. Exploratory pulling simulations on $HA2_{90}$ at different constant pulling rates (with Asp90 as the reference group for the pulling) have shown that at high rates, the coiled coil is being unfolded rather than the globular bottom. Notably, the magnitude of the pulling force, on average, does not change much in simulations with pulling rates ranging from 0.1 to 5 Å/ns. It was found that at a maximum pulling rate of 0.2 Å/ns, the globular bottom unfolds without loss of secondary structure in the coiled coil. The corresponding average pulling force of about 100 pN was therefore used in the constant force pulling simulations on $HA2_{112}$. See Appendix 4.C for more details.

The equilibration procedure for the pulling simulations on $HA2_{112}$ started with 10 000 steps of steepest descent energy minimization of each of the twenty initial configurations (obtained as described above) in vacuum. Solvent was added to the minimized structure to fill an over-sized triclinic box (12.6 nm × 6.4 nm × 5.6 nm), with the longest axis of the box aligned to the pulling direction in order to provide space for the solute to unfold. The solvated system was then energy minimized for another 10 000 steps of steepest descent. A simulation of 100 ps was then carried out in order to relax the solvent around the solute atoms, which were restrained at their initial positions. A subsequent NVT Langevin dynamics run of 50 ps then heated the system to 300 K, followed by a 100 ps NPT run for pressure equilibration. The positional restraints remained on the solute during these runs to maintain the energy minimized configuration. The resulting solvated system configuration was used as the starting configuration for a production (pulling) run. The NVT ensemble was used in the pulling simulations, because GROMACS could not handle pulling distances larger than half the box size in the NPT ensemble. During production runs, protein coordinates were saved every 1 ps for analysis.

The equilibration procedure for the equilibrium MD simulations on $HA2_{90}$ was the same, albeit that $HA2_{90}$ was solvated in a truncated dodecahedral box, sized to maintain a minimal distance of 5 Å between the protein and all periodic box walls. Also, the production runs of $HA2_{90}$ were carried out in the NPT ensemble to rule out effects from non-atmospheric pressure on the initial configurations for the pulling runs.

4.2.2 Analysis of simulation results

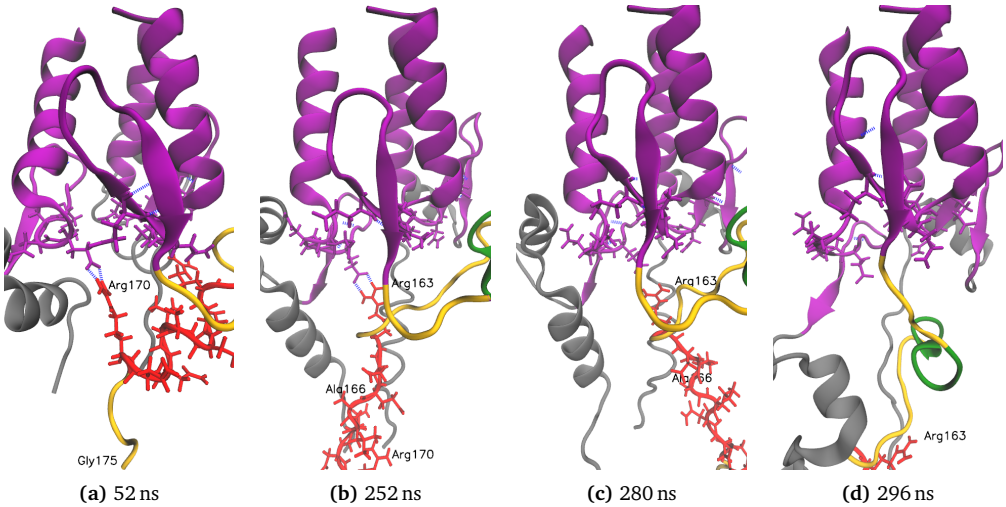
To facilitate the analysis of the results of the unfolding simulations, several segments in the globular bottom are distinguished and indicated in Figure 4.2. The N-terminal helices that belong to the coiled coil (residues 112 to 128), together with the short β-hairpin that lies on top of them (residues 129 to 145), are coloured purple in the figure and are called the 'top' part. The two alpha-helical motives in the rest of the chain are called helix G (residues 146 to 154) and helix H (163 to 171). The protein structure representations in the figures were done using VMD.[23]

In the postfusion conformation of HA2, the top part maintains its secondary and tertiary structure with respect to the prefusion structure, while the rest of the chain has unfolded and dissociated to form a leash along the central coiled coil. Therefore, the unfolding of helices G and H (loss of secondary structure) and their dissociation from the rest of the protein (loss of tertiary structure) are of primary interest. In all simulations, both helix G and H have dissociated from the top at a pulling distance of 11.5 nm between the N- and C-terminal of a chain. This distance corresponds to the length of an extended chain of residues 146 to 175. The moment at which this pulling distance was reached is called the macroscopic unfolding event, at which the simulation was stopped. For each run, we established which chain reached the macroscopic unfolding event first (the 'fastest chain'). The statistical significance of the difference between average macroscopic unfolding times was determined using a two-tailed heteroscedastic t-test, with significance levels $*p < 0.1$, $**p < 0.05$ and $***$ $p < 0.01$.

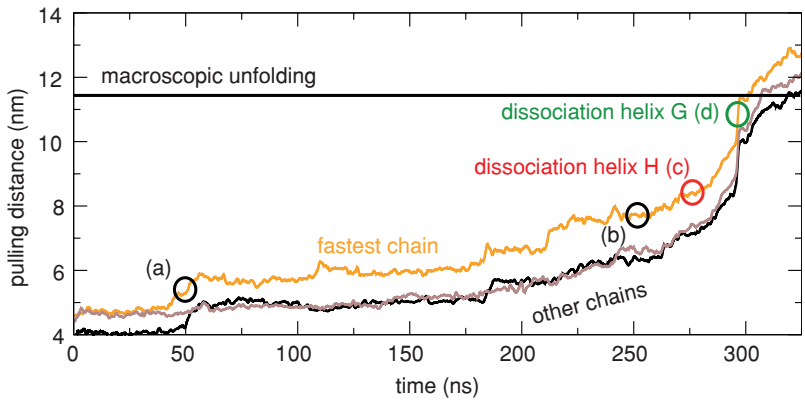
Two specific measures of time are used in describing events one length scale lower than macroscopic unfolding: the segment unfolding time and the segment dissociation time. The pulling on a chain can cause unfolding of a helical segment in the chain, but subsequent dissociation of that segment from another part of the protein might temporarily reduce the load on that segment, possibly restoring helicity. This would generally happen within 10 ns after the initial unfolding. Therefore, the unfolding time of a helix is defined as the moment at which the helical content of the segment is zero and stays zero for at least 10 ns. The helical content of a segment is taken to be the number of consecutive residues (at least three) in a helical conformation divided by the number of residues in the segment, while a residue is regarded to be in a helical conformation if its backbone dihedral angles ϕ and ψ are in the range $[-100°, -30°]$ and $[-67°, -7°]$, respectively.[15] The segment dissociation time is defined as the last moment at which the minimum distance between

all heavy atoms in the segment and all heavy atoms of the top part is less than 3.5 Å. In the calculation of the dissociation time of helix G, the top part is limited to residues 112 to 140 in order to exclude the nearest neighbours of helix G.

Hydrogen-bond analysis was employed to dissect the unfolding pathway on a residue-level. Hydrogen bonds are defined between residues with a maximum distance of 3.5 Å



(a) 52 ns **(b)** 252 ns **(c)** 280 ns **(d)** 296 ns



(e) *Evolution of chain unfolding during a single pulling simulation.*

Figure 4.3: *Series of unfolding events (a-d) in the globular bottom of H3 during a typical, concurrent unfolding pathway and corresponding pulling distance (e), including helix G and H dissociation and the macroscopic unfolding event. In (a-d) the residues in helix H of the fastest chain are shown as red sticks. The cartoon representation is used for the top (purple), the fastest chain (yellow), including helix G (green), and the other chains (gray).*

between the donor and acceptor atoms, with a 30° maximum donor-hydrogen-acceptor angle. In analogy to the segment dissociation time, a residue dissociation time, called 'last vertical contact time', is defined to be the moment of the last hydrogen bond between any residue in the range 146 to 175 and the top part. Hydrogen bond occupation was calculated as the fraction of the time that the hydrogen bond existed until the last vertical contact time.

4.3 Results

The steered molecular dynamics simulations give insight into the unfolding characteristics of the globular bottom. Our analysis of the simulations starts with the unfolding pathway, which is then used as a basis for a more in-depth study of the underlying critical interactions. First, the results of the simulations on HA from the H3N2 viral strain are presented, which will then be compared to those from the H1N1 strain.

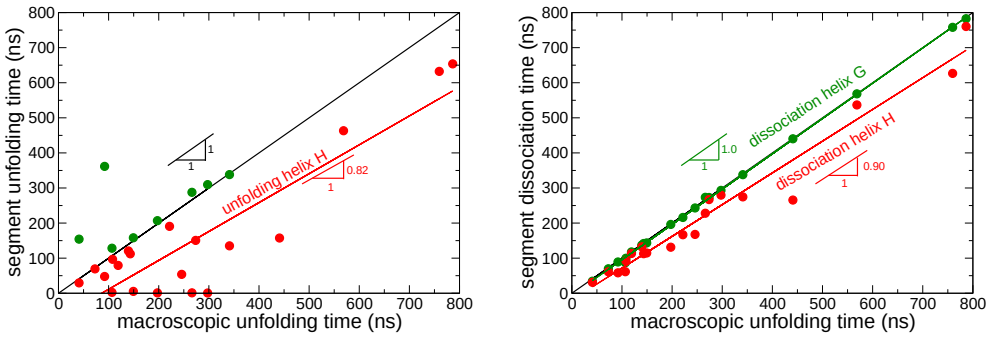
4.3.1 Unfolding pathway of the globular bottom of H3

For the globular bottom of wild-type (WT) H3, 20 steered molecular dynamics simulations were carried out, each with a different starting configuration of $HA2_{112}$. The three chains of the trimer unfolded concurrently in some simulations, while one or two of the chains clearly unfolded much faster than the other(s) in other simulations. The analysis that is presented next pertains to the fastest chain.

A typical, concurrent unfolding pathway is illustrated in Figure 4.3a to d, with a corresponding plot of the pulling distance during the simulations in Figure 4.3e. In the first 250 ns, the C-terminal of helix G (in red) is pulled away from the globular bottom (Figure 4.3a to b). While helix H is still having interactions with the top part, it unfolds (b), after which the connection between helix H and the top is lost (c). Subsequently, rapid dissociation of helix G and macroscopic unfolding ensue (d).

The correlation between the macroscopic unfolding event and loss of secondary structure in the individual segments G and H from all runs is shown in Figure 4.4a. It can be seen that in all the simulations performed, helix H unfolds prior to macroscopic unfolding. The unfolding of helix H is therefore highly correlated with macroscopic unfolding. The mean regression coefficient of 0.82 indicates that on average, helix H has unfolded at 82% of the macroscopic unfolding time. In contrast, helix G does not unfold before the macroscopic unfolding event and in most simulations not even thereafter, even though the load is still being applied. This indicates a relatively high helical stability of helix G compared to helix H, which can also be seen by the retention of secondary structure of helix G, unlike helix H, in the postfusion conformation of HA.

The dissociation of both helix G and H (loss of tertiary structure) is highly correlated with the macroscopic unfolding event. Moreover, regression coefficients of 1.0 and 0.9 for



(a) *Loss of secondary structure.* **(b)** *Loss of tertiary structure.*

Figure 4.4: *Correlation of the macroscopic unfolding time with a) unfolding times and b) dissociation times of segments helix G (green) and H (red) in the fastest chain. The dots are data points from individual simulations. The solid, coloured lines are fits from a linear regression model, with the regression coefficient indicated. The black line indicates a slope of* 1*. Helix G did not unfold in some simulations, for which some green dots and the fit are missing in the figure. The solid lines are regression curves through the data corresponding to their colour, with the regression coefficients (slopes) indicated.*

helix G and H respectively (Figure 4.4b) indicate a short delay between the dissociation of the helix and macroscopic unfolding. As soon as helix H has dissociated, macroscopic unfolding ensues rapidly and consistently. Because of this strong correlation and concurrence of helix H dissociation with the macroscopic unfolding event, we conclude that dissociation of helix H is the critical event in the unfolding of the globular bottom of HA2. We continue by unraveling the interactions between amino acids that are responsible for this critical event.

4.3.2 Identification of critical interactions

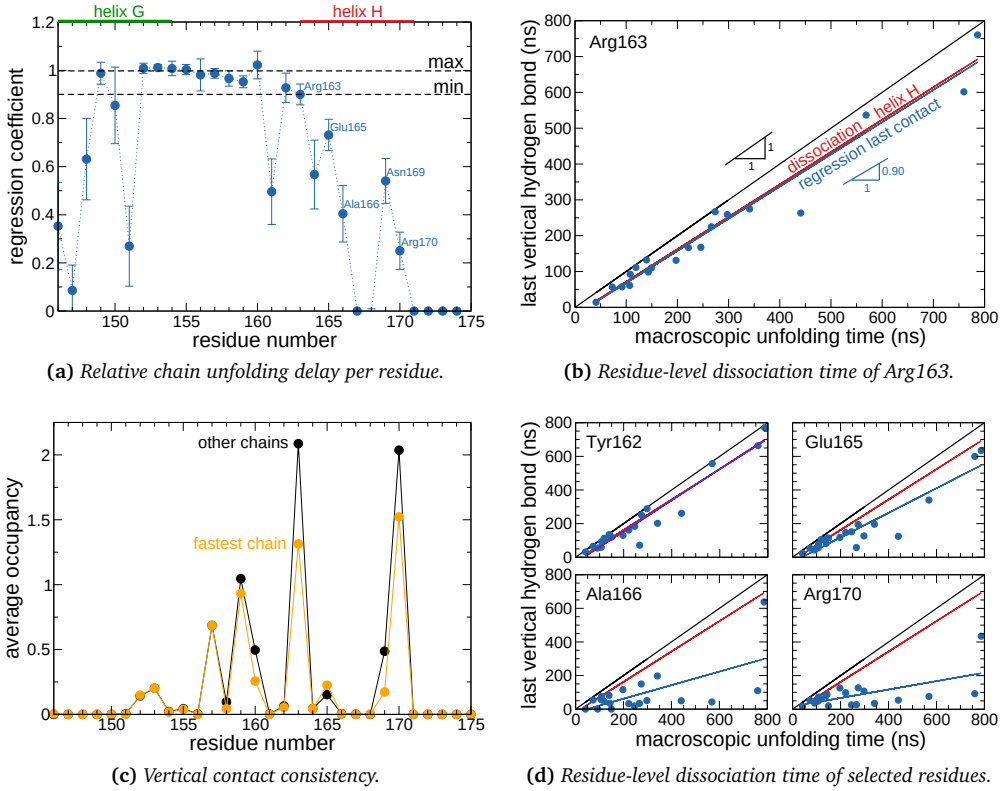
A residue-level analysis was performed to investigate which of the residues have interactions that are critical for chain stability. Since vertical contacts deliver most of the stability against the vertical pulling, we determined the 'last vertical contact time' in all simulations and studied the correlation to the time of macroscopic chain unfolding. As the chain under consideration (the fastest unfolding chain) starts to unfold from the C-terminal, residues will dissociate starting from the highest residue number (175 to smaller residue numbers). When a residue is reached that makes a vertical contact with the top, chain unfolding and dissociation will stop until the interaction is lost. A critical contact is defined as a vertical contact that significantly delays further chain unfolding and that does this consistently in (almost) all instances. Furthermore, after the critical contact has been broken, the chain should unfold without further significant delay. Hence, the amino acid closest to the C-

terminus that makes a vertical contact that consistently remains until macroscopic chain unfolding can be considered critical. The absence of further delays in chain unfolding should be reflected in a high correlation with macroscopic unfolding for that contact and a high regression coefficient, here required to be at least 0.9. Additionally, the contact should show consistency over all simulations by a low spread of the data around the regression curve. Also, to be able to provide stability at all times, critical interactions should occur throughout the simulation, so a high occupancy of vertical hydrogen bonds should be expected. Finally, it can be anticipated that the critical contact is made from within helix H, based on the previously established critical role of this segment. However, in order not to rule out important effects from residues that do not belong to helix G or H, all residues in the globular bottom were included in the current analysis.

The regression coefficient with macroscopic unfolding for the last vertical contact time per residue, and the occupancy of vertical contacts per residue are shown in Figure 4.5a and c, respectively. Starting from residue 175 on the right in Figure 4.5a, Arg170 and Arg163 have strikingly high occupancies, with on average more than two vertical hydrogen bonds per residue (Figure 4.5c). The occupancy of these residues is lower when they belong to the fastest unfolding chain, indicating that they play a significant role in stabilizing the chain. From Figure 4.5a, it can be seen that the residues Arg170 and Asn169 are the first residues that lose their vertical contacts, with regression coefficients of 0.25 and 0.54 respectively. This means that, although the occupancy of vertical hydrogen bonds made by Arg170 is high, its vertical contacts are broken relatively early in the unfolding process, as seen in Figure 4.5d. The next two residues with nonzero regression coefficients are Ala166 and Glu165, but they both have a relatively low occupancy.

The first residue that meets the requirement of a regression coefficient above 0.9 and a high occupancy is Arg163. The consistency of vertical contacts made by Arg163, the third requirement, can be seen from the relatively small error in the regression coefficient for this residue. By looking at the individual correlation data of Arg163 as shown in Figure 4.5b, it can indeed be seen that the data points have a low spread around the regression curve (blue). Furthermore, this curve almost coincides with the regression curve of helix H dissociation (red), highlighting the strong interconnection between the loss of vertical contacts by Arg163 and the dissociation of helix H.

The next residue in the chain, Tyr162, also has a high regression coefficient, but the data is more spread out (Figure 4.5d). Also, as the occupancy of vertical contacts made by Tyr162 is very low, it can be inferred that the stability provided by Arg163 allows Tyr162 to make incidental vertical contacts until the contacts made by Arg163 are broken and helix H dissociates. Altogether, its large regression coefficient, high occupancy and high consistency of vertical contacts convincingly identify Arg163 to be the critical amino acid.



(a) *Relative chain unfolding delay per residue.*

(b) *Residue-level dissociation time of Arg163.*

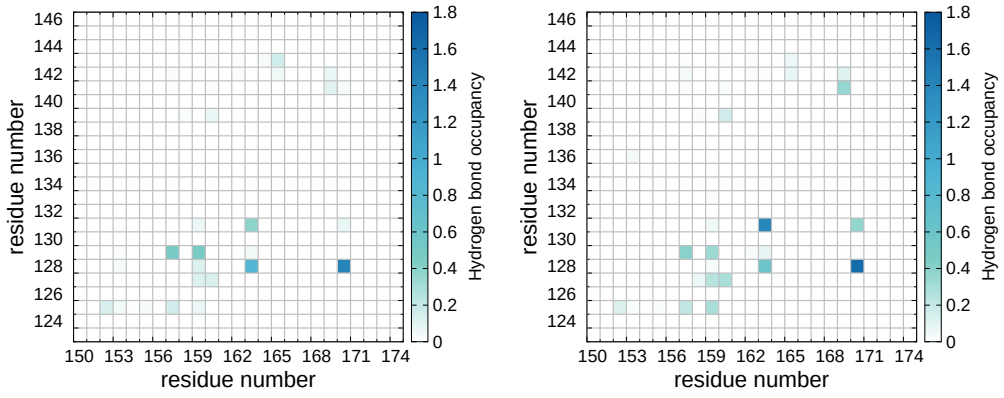
(c) *Vertical contact consistency.*

(d) *Residue-level dissociation time of selected residues.*

Figure 4.5: *Residue-level analysis of unfolding based on the last vertical hydrogen bond between a residue in the unfolding chain and the top of the protein domain. a) Regression coefficients, obtained from the correlation between the last vertical contacts in the fastest chain and macroscopic unfolding time as shown, e.g., for Arg163 in b). c) Vertical hydrogen-bond occupancy per residue until the last vertical contact time for the fastest chain and the 'other chains', averaged over 20 unfolding simulations of WT H3. The error bars are smaller than the symbols. d) Individual correlation plots of the last vertical contact time with the macroscopic unfolding time for residues Tyr162, Glu165, Ala166 and Arg170.*

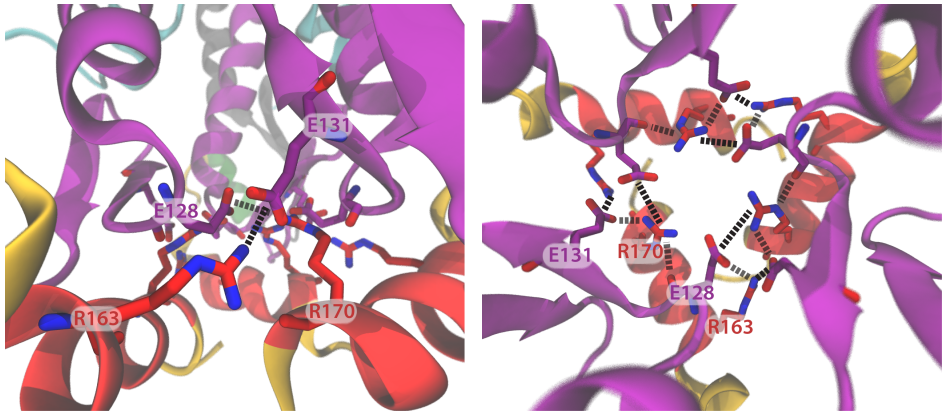
4.3.3 Salt-bridge network

Having identified the residue in the globular bottom that makes critical interactions with the top, we are also interested in which residues in the top participate in these critical interactions. As such, the hydrogen-bond occupancy between the residues in the bottom and in the top is mapped in Figure 4.6. Together with Arg163, the residues Glu128, Glu131 and Arg170 all show a high occupancy in these contact maps. The residues Glu128 and Glu131 are positioned at the C-terminal of the central coiled coil and in the β-strand that doubles back along the coiled coil, respectively. At neutral pH, these glutamates are neg-



(a) *Fastest chain contact map.*

(b) *Other chains contact map.*



(c) *Side view on the salt-bridge network.*

(d) *Top view.*

Figure 4.6: *a) Hydrogen bond contact map of the vertical contacts made by the fastest unfolding chain and b) the 'other chains'. The colour indicates the occupancy of vertical contacts for that residue pair until the last vertical contact time, averaged over all simulations of WT H3. c) Side and d) top view of the prefusion crystal structure of WT H3, showing the salt-bridge network in the core of the globular bottom. The segment colouring corresponds to that of Figure 4.1. Glu128, Glu131, Arg163 and Arg170 are drawn in licorice representation, labeled at their α-carbon atom and coloured as the corresponding segment. The black dashed lines indicate the four most occupied vertical hydrogen bonds.*

atively charged and are positioned in between the positively charged Arg163 and Arg170 in the prefusion crystal structure (see Figure 4.6a and b). The proximity and nature of these oppositely charged residues, experiencing both electrostatic attraction and hydrogen bonding, allow them to form a network of stabilizing salt bridges in the core of the globular bottom.

In the fastest unfolding chain (Figure 4.6c), Arg163 prefers to interact with Glu128 but also has contact with Glu131, while Arg170 predominantly has interactions with Glu128. In the other chains (Figure 4.6d), both Arg163 and Arg170 have interactions with either of the glutamates, while Arg163 prefers Glu131 and Arg170 prefers Glu128. The generic sequence of events that can be deduced from this contact map and a close look at the individual trajectories is as follows. In the fastest unfolding chain, Arg170 is being pulled away from Glu131 rather early in the unfolding process, and will only make contact to Glu128 until their moment of last contact at, on average, around 25% of the macroscopic unfolding time (see Figure 4.5a and d). Meanwhile, in the same chain, Arg163 remains in contact with its preferred partner Glu131, until this contact breaks. The remaining contact of Arg163 with Glu128 lies beneath Glu131, and therefore causes a significant delay in the macroscopic unfolding event. Thus, the dominant role of Arg163 in the stability of the globular bottom resides in its ability to establish two stabilizing vertical interactions during unfolding.

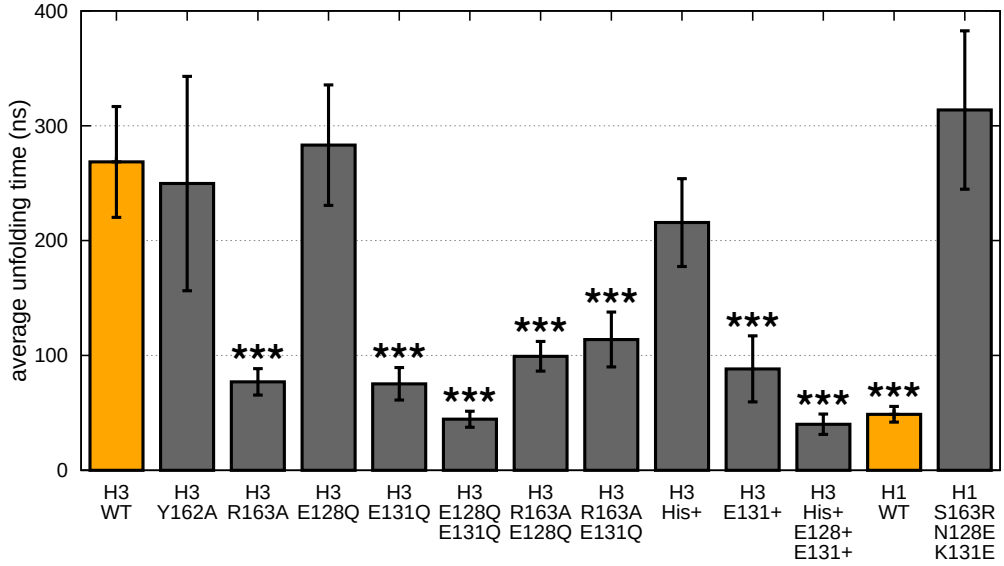


Figure 4.7: *Average macroscopic unfolding times from constant force pulling simulations on the WT (orange) and mutated (gray) globular bottom of H3 and H1 at neutral pH, and WT H3 in different protonation states. The error bars indicate the standard error of the mean. The stars indicate the significance level for the difference of the average with respect to the H3 WT (*** $p < 0.01$). A table with significance levels between all given averages can be found in Appendix 4.D, Table 4.D.1.*

4.3.4 Mutation studies

To corroborate the influence of specific residues on the unfolding rate of the globular bottom, a mutation study was performed. For each mutation, 10 to 20 simulations (depending on the variance of the macroscopic unfolding time) were performed with settings and starting configurations identical to the WT simulations (except for the mutation). The change in stability due to the mutation is inferred via the average macroscopic unfolding time of the mutant compared to the WT, as shown in Figure 4.7. The macroscopic unfolding time of the WT was 269 ± 48 ns on average.

To confirm the decisiveness of vertical contacts made by Arg163, it was mutated to an alanine residue (R163A). Alanine has a high helical propensity, so that the α-helical structure of helix H is not disrupted by the mutation, but the ability of Arg163 to form salt bridges is diminished. The average macroscopic unfolding time for the R163A mutant was 77 ± 12 ns (see Figure 4.7), significantly smaller than that of the WT and clearly validating the designation of R163A as a critical residue. Mutation of Tyr162, a direct neighbour of Arg163, resulted in an insignificant difference in the macroscopic unfolding time with respect to the WT (see Figure 4.7), confirming the earlier suggestion that the high regression coefficient for the last vertical contact times of Tyr162 arises just because of the stability offered by Arg163.

Next, the salt bridging partners of Arg163 were mutated. Although Arg163 makes frequent contacts with Glu128 in the WT (see Figure 4.6c and d), the E128Q mutation with uncharged glutamine did not have a significant effect on the macroscopic unfolding time. Detailed analysis of the results revealed that the long unfolding times of E128Q are caused by the additional stabilization from a salt bridge formed between Asp160 and Arg127 that was prohibited by the presence of Glu128 in the WT. The significance of the salt bridge between Arg163 and Glu131 is once more confirmed by the E131Q mutation, resulting in an almost identical average macroscopic unfolding time as the R163A mutant. The double mutation E128Q-E131Q, in turn, results in significantly faster unfolding than only the E131Q mutation, as it takes away all the opportunities of Arg163 and Arg170 to establish vertical contacts. The other two double mutants of H3, R163A-E128Q and R163A-E131Q, show a slight increase in the average macroscopic unfolding time with respect to the R163A mutant. However, the difference is not significant (see Table 4.D.1 in Appendix 4.D), while they still unfold much faster than the WT. This is to be expected because due to the R163A mutation, neither Glu128 nor Glu131 is involved in a critical vertical contact anymore.

4.3.5 Protonation studies

The simulations discussed so far, were performed at pH 7. However, as the pH of HA activation lies between pH 5 and 6, there are a number of residues that would get protonated during the rearrangements. Therefore, the unfolding behaviour of the globular bottom

at the pH of HA activation was investigated using static protonation of specific residues. Histidine, with a model pKa of 6.5, would be the first to be protonated. So, initially, a simulation with all histidine residues in our truncated model ($HA2_{112}$) protonated was performed (called His+). As shown in Figure 4.7, this did not lead to a significant change in the unfolding time of the globular bottom, in line with the observation that none of the histidines were involved in a critical vertical contact. However, the glutamic acid residues Glu128 and Glu131, with a model pKa of 4.15, could undergo a structure-dependent pKa shift and obtain a proton around pH 5. Upon protonation, the electrostatic interaction between the arginine and glutamic acid residues would become disrupted and each glutamic acid would lose a hydrogen-bond acceptor. Indeed, protonation of Glu131 (E131+) leads to an average unfolding time close to that of the E131Q mutant. Similarly, the combined protonation of all histidines, Glu128 and Glu131 (His+E128+E131+) is in line with the results of the E128Q-E131Q mutant.

4.3.6 Conservation of stabilizing amino acids

Having shown the effect of mutation or protonation within the salt-bridge network in H3, conservation of these residues would implicate them in a generic, conserved molecular

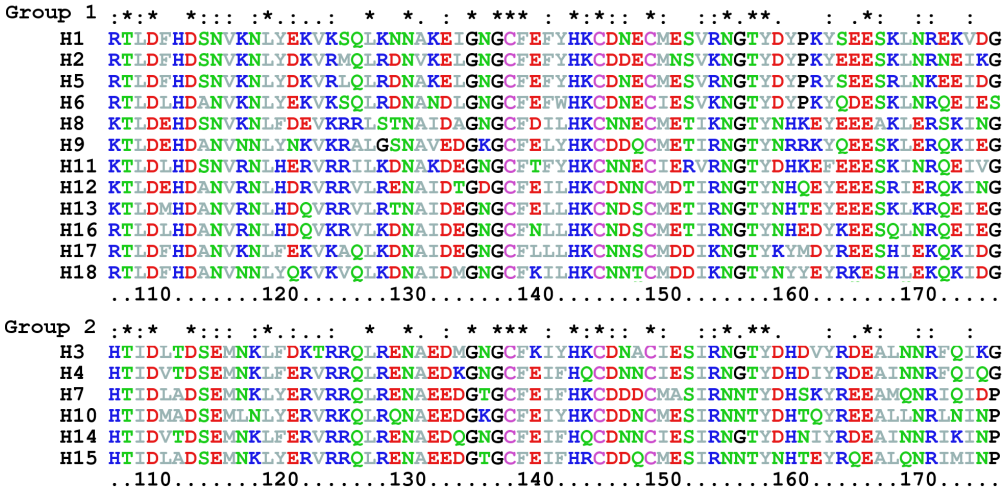


Figure 4.8: *ClustalW*[29] *multiple sequence alignment of residues 106 to 175 from HA2 of HA subtypes H1 to H18, organized by phylogenetic group. The sequences were obtained from the online Influenza Virus Resource,* [1] *taken from the strains listed in Table 4.D.2 in Appendix 4.D. The residues are coloured depending on their type at pH 5: hydrophobic (gray), cysteine (pink), uncharged polar (green), positively charged (blue) and negatively charged (red). Residues that are fully conserved over all subtypes are indicated with a '*' on top, while mutations that maintain the same biochemical properties are denoted with a ':' (conservative mutation) or a '.' (semi-conservative mutation).*

mechanism for the regulation of the stability of the globular bottom throughout HA subtypes. To explore this, a sequence alignment for the globular bottom domain of all 18 HA subtypes is shown in Figure 4.8. Apart from H1 (PR8) and H3 (X31), which were selected because there is data on the fraction of non-productive HAs for these strains, the sequences compared here were taken from arbitrary influenza virus strains, as listed in Table 4.D.2 in Appendix 4.D. HA subtypes are divided into two groups, based on their phylogenetic classification. Group 1 is characterized by the histidine at position 111. This histidine is located at position 106 in group 2, to which H3 belongs. It can be seen in Figure 4.8 that within this group, Glu128, Glu131, Arg163 and Arg170 are fully conserved, except for one conservative mutation (E128Q) in H10.

Unlike in group 2, the salt-bridge network is not conserved within group 1. The critical residue Arg163 is only present in two subtypes in group 1 (H17 and H18) and has mutations to other polar residues otherwise. Although Arg170 is conserved as a basic residue throughout both groups, the glutamates that act as hydrogen-bond partners for the asparagines in the network are only scarcely found in group 1. There are mutations to other polar residues at position 128, as well as hydrophobic and even positively charged residues at position 131. Because most of the residues in the salt-bridge network are absent in group 1, the unfolding behaviour of the globular bottom from a subtype in group 1 might be different from that of H3 studied so far. This will be investigated in the next section.

4.3.7 Stability of the globular bottom of H1

The stability of the globular bottom of H1 was studied using a setup identical to the pulling simulations on H3. From the subtypes in group 1, H1 was selected because of the availability of experimental data.[36] Although the pulling characteristics of H1 could be different from those of H3, which would require adjustment of the pulling force, the same pulling force of 100 pN was used to enable a direct comparison of the results. As shown in Figure 4.7, the resulting average macroscopic unfolding time of WT H1 is 49 ± 7 ns, which is significantly smaller than that of WT H3 and is more comparable to the E128Q-E131Q mutant.

We next tried to stabilize the globular bottom of H1, informed by the results on H3. In the sequence alignment in Figure 4.8, the critical arginine of H3 at position 163 is replaced with a serine in H1, but applying an R163S mutation did not increase the average macroscopic unfolding time of H1 in our simulations (data not shown). This can be attributed to a lack of salt bridging partners for Arg163 in the top of H1. With the triple mutation S163R-N128E-K131E however, all residues participating in the salt-bridge network of H3 were present. This increased the average unfolding time of H1 by more than a factor 6, to 314 ± 69 ns (Figure 4.7), which is not significantly different from that of WT H3, showing that the salt-bridge network of H3 plays an important role in providing mechanical

stability to the globular bottom.

4.3.8 Single-particle experiments

The predictions from the pulling simulations have been tested in single-particle fusion kinetics experiments by colleagues in the Van Oijen lab at the University of Groningen, The Netherlands, similar to the work by Otterstrom et al.[36] The details of the assay, as described by our colleagues, can be found in Appendix 4.A. In short, fluorescence microscopy was used to track individual virus particles, that fuse with a supported lipid bilayer. Labeled antibodies were used that bind to HA and inhibit its conformational changes, thus mimicking the effect that a non-productive HA would have. The amount of antibodies per viral particle was estimated from their fluorescence intensity. The yield of fusion, defined as the fraction of the total number of particles for which hemifusion was detected, was determined for different numbers of these inhibitors per virus particle.

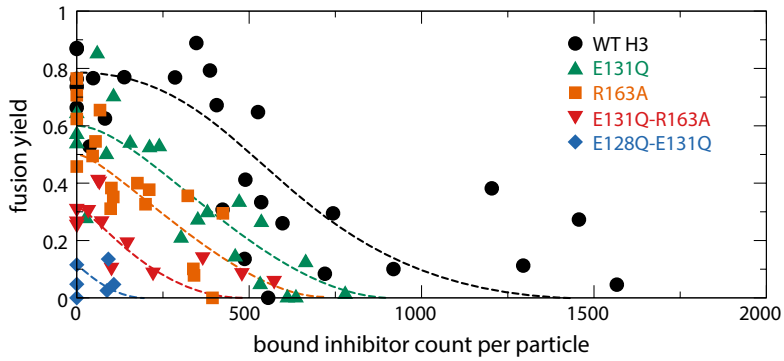


Figure 4.9: *Stoichiometry of fusion inhibition and antibody binding from single-particle experiments as supplied by the Van Oijen lab.[6] Data is shown for WT H3 (black circles) and the mutants E131Q (green upwards triangles), R163A (orange squares), E131Q-R163A (red downwards triangles) and E128Q-E131Q (blue diamonds). The coloured dashed lines are drawn as a guide to the eye through the correspondingly coloured data points for each mutant, inspired by the shape of the curves from the stochastic model of Ivanovic et al.[25]*

The results of the experiments are shown in Figure 4.9. Indicative curves (dashed) are drawn as a guide to the eye through each of the data sets, as a schematic representation of the stochastic model of Ivanovic et al.[25] The results on WT H3 are in agreement with those reported by Otterstrom et al.[36] In the first of three regimes that can be identified, the fusion yield remains constant at about 0.8 (below 500 inhibitors per particle) because there are still enough HAs left to mediate fusion, despite the inhibition of a number of them by the antibodies. In the second regime (between 500 and 1500 inhibitors), not all of the particles have enough productive HAs left to mediate fusion and the fusion yield

starts to decrease. For WT H3, the half-maximum fusion yield that can be used to quantify the location of this transition regime lies at about 600 inhibitors per particle, as estimated from the indicative dashed line. In the third regime (above 1500 inhibitors), the fusion yield goes to zero because all available HAs have been inhibited.

The stoichiometry of fusion inhibition was also determined for a number of the H3 mutants described in Section 4.3.4. The mutations that destabilize the globular bottom relative to the WT in the simulations are hypothesized to have a higher fraction of non-productive HAs on the virus particle. As can be seen in Figure 4.9, the estimated half-maximum fusion yield is decreased from 600 to 400 inhibitors by the E131Q mutation. This indicates that the E131Q mutant virus has a higher fraction of non-productive HAs than the WT, possibly because of the destabilization of the globular bottom induced by knocking out the salt bridge between Glu131 and Arg163. The R163A mutation shifts the half-maximum yield further to the left (to an estimated 300 inhibitors) compared to the E131Q mutation. There is no quantitative difference between the average unfolding times of the E131Q and R163A mutations in the simulations (Figure 4.7) that could explain this difference in observations. However, the contact maps that were discussed in Section 4.3.3 point towards an added destabilizing effect of the R163A mutation, by knocking out the vertical hydrogen bonds between Arg163 and both Glu128 and Glu131 at once. The double mutation E131Q-R163A decreases the half-maximum fusion yield even further (to about 200 inhibitors), presumably by leaving only the interaction between Glu128 and Arg170 in the salt-bridge network. Finally, the fusion yield of the E128Q-E131Q mutation is 0.15 at most. Based on our simulations, this result might be attributed to the fact that the double mutation knocks out all salt bridging partners in the top and effectively destroys the network.

4.4 Discussion

The commonly accepted productive pathway of HA-mediated membrane fusion (top row in Figure 4.1) is based on the assumption that the rate of B-loop extension is higher than the rate of collapse in the C-terminal part of HA2. Recently however, it was found that a major fraction of HA refolds non-productively during the fusion process,[25] for reasons that are not clear at this stage. Here, we tested the hypothesis that destabilization of the globular bottom of HA2 can reduce the productivity of HA by favouring the non-productive refolding pathway (bottom row in Figure 4.1), as hinted at by Lin et al.[30]

Using steered molecular dynamics simulations to investigate stabilizing molecular interactions in the globular bottom of HA, Arg163 was identified as a critical amino acid, which is part of a network of salt bridges in the core of the globular bottom that controls its stability. This salt-bridge network appears to coincide with some of the critical interactions for globular bottom unfolding that were found in simulations of HA by Lin et al.[30] in the static prefusion crystal structure of H3. We here propose that Arg163 also

belongs to this cluster because of its interactions with Glu128 and Glu131, as indicated by our simulations. Ni et al.[35] have identified a second cluster, Glu150, Arg153 and His26, which lies on the HA2-HA2 interface between the globular bottom and the fusion peptide. However, the fusion peptide is the first to dissociate after HA activation[16] and does not play any further role in the collapse of the globular bottom. For this reason, we did not include His26 in our study. The third cluster reported by Ni et al.[35] consists of Asp158, His159 and Asp160 in the globular bottom and Arg127 and Asn129 in the top of H3. Interactions between these residues appeared clearly in the analysis of our simulations of H3 (Figures 4.5c and 4.6), albeit with less occupancy and with a less dominant role than the residues in the first cluster. In our simulations of H1 however, the residues corresponding to this third cluster do play a significant role. In particular, Tyr157, Asp158 and Tyr159 in the globular bottom of H1 form hydrogen bonds with Asn128 and Asn129, and have hydrophobic π-stacking interactions with Tyr141 and His142 (data not shown). Of these, residues Asn129 and Tyr157 are fully conserved among HA subtypes (see Figure 4.8). Our results further show that in absence of the salt-bridge network residues Glu128, Glu131 and Arg163, the globular bottom of H1 is significantly less stable than that of H3. We hypothesize that this relative instability causes the higher fraction of non-productive HAs observed for H1.[25] Conversely, it might be possible to stabilize the globular bottom of H1 by reintroducing these residues, as in our simulations of the S163R-N128E-K131E mutant, for which experimental verification is in progress.

Besides the influence of mutations on the pathway of conformational changes of HA discussed so far, our results suggest that protonation of residues in the salt-bridge network might act as an intermediate trigger for the collapse of the globular bottom. HA activation has been shown to start with the release of the fusion peptide below pH 6,[24] corresponding to the acidity inside the maturing-to-late endosome.[28] At this point, the stability of the salt-bridge network in the globular bottom gives the fusion peptides time to insert into the membrane, because the participating glutamates are not expected to be protonated yet (based on their model pKa of 4.2). As the endosome matures, more HAs become activated and insert into the membrane, while the pH drops further. Below a certain pH, Glu128 and Glu131 do become protonated and initiate the collapse of the globular bottom and subsequent zippering. This could happen at a pH above 4.2, because the local electrostatic environment within the protein can greatly shift the pKa of a residue with respect to its model pKa.[17,48] The existence of such an intermediate trigger could explain the high fraction of non-productive HAs in the experiments, in which the instantaneous drop to pH 5 might trigger globular bottom collapse concurrently with fusion peptide release. In contrast, the pH drop is more gradual *in vivo*, so HA would have more time to insert its fusion peptides into the membrane before a further decrease in pH triggers globular bottom collapse, resulting in a higher productive fraction. Comparing the fusion yield and the fraction of non-productive HAs at different levels of pH could be used to investigate this hypothesis further.

There might be another role for protonation of selective residues, in causing symmetry-breaking of the extended intermediate. Several authors have already commented on the topological difficulties when the threefold-symmetric HA is rearranging from the extended intermediate to the postfusion conformation, while being attached to both the viral and the target membrane.[2,18,30] The asymmetry that is thus needed can arise due to stochastic events in the timescale of globular bottom unfolding, such as the unsynchronized switching of molecular interactions within individual chains. As an example, the protonation of the buried residues Glu128 and Glu131 can switch off a salt-bridge interaction and is a highly structure-dependent, stochastic event. So, it is not likely that all of these residues will be protonated in all the chains at once. Our simulations indicate that as soon as one of these glutamic acids becomes protonated in a certain chain, the dissociation of this chain would cause consolidation of the other chains. The already unfolded chain could then proceed to the zippering, thereby facilitating the interruption of symmetry. This symmetry-broken state would allow HA to position itself more parallel to the membranes, making room for the membranes to approach each other. This hypothesized asymmetric unfolding of the globular bottom could be explored with long-timescale equilibrium simulations or further pulling simulations, with only selected residues within a single chain protonated or mutated.

There are a number of assumptions, limitations and possible shortcomings within our model that should be addressed. First of all, our current hypothesis assumes the existence of an extended intermediate, for which there is still no direct evidence. However, an alternative pathway of HA rearrangements that does not include the formation of an extended intermediate also relies on a low unfolding rate of the globular bottom, as proposed by Lin et al.[30] In this pathway, HA activation causes destabilization within the hinge region (residues 107 to 112), around which the protein bends while the globular bottom remains intact. The fusion peptides that have been released due to the destabilization can then diffuse into the viral or the target membrane. With at least one of the fusion peptides in the target membrane, unfolding of the globular bottom allows the combined action of coiled-coil extension and the zippering mechanism to pull the membranes together. Hence, in this alternative pathway, destabilizing the globular bottom would also decrease the window of opportunity for one or more fusion peptides to insert into the target membrane and would therefore decrease HA productivity.

In our simulations, artifacts might have been introduced by disregarding all of HA1 and part of HA2 in our simulations. However, at the moment that globular bottom stability becomes relevant in the pathway of HA rearrangements, the fusion peptide has already dissociated and there are only a few interactions between the outside of the globular bottom and HA1. So, while the truncation into $HA2_{112}$ was mainly done to make the simulations computationally more feasible, we assume that all the relevant interactions within the globular bottom itself were still included. Further, the pulling conditions that were applied could have biased the results. However, the direction of pulling parallel to the long

axis of the central coiled coil is reminiscent of the vertical load that would be applied to an extended and inserted intermediate by membrane undulations in the biological context. Also, although the pulling force of 100 pN was much higher than it would be in the biological context, it was carefully calibrated to still allow for stochasticity in intramolecular interactions. Furthermore, as evident from our results, the stochastic interruption of a number of critical contacts determines the unfolding rate. So, although the used pulling direction and force would not necessarily lead to the same unfolding behaviour as *in vivo*, we are confident that the simulations are qualitatively correct in identifying interactions between residues that are critical for the stability of this domain of HA under the applied loading. We assume that these residues will be important in other unfolding pathways as well. We also note that, due to the temporal spread of the hydrogen bonds breaking, the resulting macroscopic unfolding times showed a large variance. However, 20 simulations were carried out for the relatively stable globular bottoms that showed a large variance in macroscopic unfolding times and 10 for the less stable ones that displayed much lower variance. This was sufficient to calculate the significance levels between the averages and obtain statistically relevant results, with standard errors of the mean within about 20 % of the average.

A further limitation of the currently presented simulation results is the focus on vertical contacts. Although we did not include them in our results, horizontal contacts between the chains within the globular bottom did play a role in unfolding. The collection of critical contacts that have been reported by Lin et al. [30] also contained such horizontal interactions. However, due to the asynchronous unfolding of individual chains in our simulations, the residues in one chain slide along the residues of the neighbouring chain during the pulling. This results in inter-chain interactions between almost all possible residue pairs, creating a smeared out contact map. Because we did not find any consistent, critical interactions within this data, we have chosen to focus only on vertical contacts made by the fastest unfolding chain. A more complete profile of inter-chain interactions would be possible by enforcing concurrent unfolding of the three chains, as if they were all connected to the same plane (such as a membrane).

In the preliminary experimental results, decreased fusion efficiency was observed for the R163A mutation, as well as for other disrupting mutations within the salt-bridge network. This is consistent with the hypothesis that the stability of the globular bottom influences the number of productively rearranging HA proteins on the virus particle. However, there could be alternative explanations for the trends observed in the experiments. Firstly, any of the mutations could have led to differences in HA folding, HA expression or the integrity of the virus particle as a whole during their production. For example, increased particle size or lower HA incorporation on the virus particle would effectively reduce the number of productive HAs as well. So, the mutant viruses should be examined for these effects in a future experiment. Secondly, the mutation in the globular bottom could have increased or decreased the activation pH of HA, e.g., by destabilizing the fusion peptide. In

fact, a study by Daniels et al.[13] has shown an elevated pH of H3 activation by a double mutation that includes Arg163. However, with such an increase in triggering pH, a reversed trend would be expected in the experiments. After all, more HAs would be activated at the same pH, effectively increasing the fraction of HAs participating in the fusion process. Conversely, a decrease in activation pH would lead to the trend as observed in the experiments indeed. Although there are no direct indications in the prefusion conformation that would explain such a stabilization of the fusion peptide by one of the mutations, computer simulations have already shown that globular bottom unfolding can induce fusion peptide release.[31] So, to definitively rule out any effect of the mutations on HA activation, the activation pH of the mutants should be measured. Lastly, mutation of Arg163 could cause the observed decrease in fusion efficiency by a destabilizing effect on the zippering mechanism. Because Arg163 is part of the loop that binds onto the coiled coil in the postfusion structure, the R163A mutation could reduce the binding affinity of this loop and thereby the energy that HA can provide in the fusion process. Hence, more HAs would be needed to mediate fusion, leading to the observed trend. Measuring the thermodynamic stability of the postfusion conformation of the R163A mutant, to compare with that of the WT,[10] could give more information on the extent of this effect.

4.5 Conclusion

Our mechanical unfolding simulations of the globular bottom of HA suggest a pivotal stabilizing role for four ionic residues: Glu128, Glu131, Arg170 and, in particular, the critical residue Arg163. The salt-bridge network that these residues form is conserved throughout phylogenetic group 2 of influenza A, but the participating residues are mostly absent in group 1. The observed trends in preliminary single-particle fusion kinetics experiments might be consistent with a model of HA-mediated fusion in which destabilizing the globular bottom causes an increased fraction of non-productively refolding HAs on the virus surface. However, further experiments are necessary to rule out that other effects from the mutations are causing the observed trends. Knowing the specific, conserved amino acids that determine the productivity of HA can potentially aid in the development of therapeutics, using a strategy that targets conserved residues that are critical for protein function.

References

[1] Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, et al. 2008. The influenza virus resource at the National Center for Biotechnology Information. *J. Virol.* 82:596–601

[2] Baquero E, Albertini AA, Vachette P, Lepault J, Bressanelli S, Gaudin Y. 2013. Intermediate conformations during viral fusion glycoprotein structural transition. *Curr. Opin. Virol.* 3:143–150

[3] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, et al. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273

[4] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, et al. 2014. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucl. Acids Research* 42:W252–W258

[5] Blijleven JS, Boonstra S, Onck PR, van der Giessen E, van Oijen AM. 2016. Mechanisms of influenza viral membrane fusion. *Semin. Cell Dev. Biol.* 60:78–88

[6] Blijleven JS, van Oijen AM. 2017. *Private communication*

[7] Boonstra S, Onck PR, van der Giessen E. 2016. CHARMM TIP3P water model suppresses peptide folding by solvating the unfolded state. *J. Phys. Chem. B* 120:3692–3698

[8] Bullough PA, Hughson FM, Skehel JJ, Wiley DC. 1994. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371:37–43

[9] Carr CM, Kim PS. 1993. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73:823–832

[10] Chen J, Skehel JJ, Wiley DC. 1999. N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. *PNAS* 96:8967–8972

[11] Chernomordik LV, Kozlov MM. 2003. Protein-lipid interplay in fusion and fission of biological membranes. *Annu. Rev. Biochem.* 72:175–207

[12] Costello DA, Whittaker GR, Daniel S. 2015. Variations in pH sensitivity, acid stability, and fusogenicity of three influenza virus H3 subtypes. *J. Virol.* 89:350–360

[13] Daniels RS, Downie JC, Hay AJ, Knossow M, Skehel JJ, et al. 1985. Fusion mutants of the influenza virus hemagglutinin glycoprotein. *Cell* 40:431–439

[14] Di Lella S, Herrmann A, Mair CM. 2016. Modulation of the pH stability of influenza virus hemagglutinin: A host cell adaptation strategy. *Biophys. J.* 110:2293–301

[15] García A, Sanbonmatsu K. 2002. α-Helical stabilization by side chain shielding of backbone hydrogen bonds. *PNAS* 99:2782–2787

[16] Garcia NK, Guttman M, Ebner JL, Lee KK. 2015. Dynamic changes during acid-induced activation of influenza hemagglutinin. *Structure* 23:665–676

[17] Harms MJ, Castaneda CA, Schlessman JL, Sue GR, Isom DG, et al. 2009. The pK(a) values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol.* 389:34–47

[18] Harrison SC. 2008. Viral membrane fusion. *Nat. Struct. Mol. Biol.* 15:690–698

[19] Harrison SC. 2015. Viral membrane fusion. *Virology* 479:498–507

[20] Hess B. 2008. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* 4:116–122

[21] Hess B, Kutzner C, van der Spoel D, Lindahl E. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447

[22] Hu J, Rapoport TA. 2016. Fusion of the endoplasmic reticulum by membrane-bound GTPases. *Semin. Cell Dev. Biol.* 60:105–111

[23] Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* 14:33–38

[24] Ivanovic T, Choi JL, Whelan SPJ, van Oijen AM, Harrison SC. 2013. Influenza virus membrane fusion by cooperative fold-back of stochastically induced hemagglutinin intermediates. *eLife* 2:e00333

[25] Ivanovic T, Harrison SC. 2015. Distinct functional determinants of influenza hemagglutinin-mediated membrane fusion. *eLife* 4:e11009

[26] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935

[27] Kielian M. 2014. Mechanisms of virus membrane fusion proteins. *Annu. Rev. Virol.* 1:171–189

[28] Lagache T, Sieben C, Meyer T, Herrmann A, Holcman D. 2017. Stochastic model of acidification, activation of hemagglutinin and escape of influenza viruses from an endosome. *Front. Phys.* 5:25

[29] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. 2007. Clustal W and clustal X version 2.0. *Bioinformatics* 23:2947–2948

[30] Lin X, Eddy NR, Noel JK, Whitford PC, Wang Q, et al. 2014. Order and disorder control the functional rearrangement of influenza hemagglutinin. *PNAS* 111:12049–12054

[31] Lin X, Noel JK, Wang Q, Ma J, Onuchic JN. 2016. Lowered pH leads to fusion peptide release and a highly dynamic intermediate of influenza hemagglutinin. *J. Phys. Chem. B* 120:9654–9660

[32] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586–3616

[33] Martens S, McMahon HT. 2008. Mechanisms of membrane fusion: Disparate players and common principles. *Nat. Rev. Mol. Cell Biol.* 9:543–556

[34] Miyamoto S, Kollman PA. 1992. Settle - An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962

[35] Ni F, Chen X, Shen J, Wang Q. 2014. Structural insights into the membrane fusion mechanism mediated by influenza virus hemagglutinin. *Biochemistry* 53:846–854

[36] Otterstrom JJ, Brandenburg B, Koldijk MH, Juraszek J, Tang C, et al. 2014. Relating influenza virus membrane fusion kinetics to stoichiometry of neutralizing antibodies at the single-particle level. *PNAS* 111:E5143–E5148

[37] Park HE, Gruenke JA, White JM. 2003. Leash in the groove mechanism of membrane fusion. *Nat. Struct. Biol.* 10:1048–1053

[38] Rico F, Gonzalez L, Casuso I, Puig-Vidal M, Scheuring S. 2013. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science* 342:741–743

[39] Rizo J, Xu J. 2015. The synaptic vesicle release machinery. *Annu. Rev. Biophys.* 44:339–367

[40] Sauter NK, Hanson JE, Glick GD, Brown JH, Crowther RL, et al. 1992. Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: Analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. *Biochemistry* 31:9609–9621

[41] Schelker M, Mair CM, Jolmes F, Welke RW, Klipp E, et al. 2016. Viral RNA degradation and diffusion act as a bottleneck for the influenza A virus infection efficiency. *PLOS Comput. Biol.* 12:e1005075

[42] Schrödinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.6

[43] Skehel JJ, Wiley DC. 1998. Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* 95:871–874

[44] Stegmann T, Delfino JM, Richards FM, Helenius A. 1991. The HA2 subunit of influenza hemagglutinin inserts into the target membrane prior to fusion. *J. Biol. Chem.* 266:18404–18410

[45] Weber T, Paesold G, Galli C, Mischler R, Semenza G, Brunner J. 1994. Evidence for H+-induced insertion of influenza hemagglutinin HA2 N-terminal segment into viral membrane. *J. Biol. Chem.* 269:18353–18358

[46] WHO. A revision of the system of nomenclature for influenza viruses: a WHO Memorandum. 1980. *Bull. World Health Organ.* 58:585–591

[47] Wilson IA, Skehel JJ, Wiley DC. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289:366–373

[48] Zhou Y, Wu C, Zhao L, Huang N. 2014. Exploring the early stages of the pH-induced conformational change of influenza hemagglutinin. *Proteins* 82:2412–2428

4.A Appendix: Single-particle assay

The experiments were performed to match the work by Otterstrom et al.[36] as closely as possible, and a more detailed explanation of the methods can be found there. A brief overview follows. Materials were purchased from Sigma-Aldrich unless noted otherwise.[†]

Virus and R18 labeling

Influenza A strains A/Aichi/2/1968 (X31, H3N2) and A/Puerto Rico/8/1934 (PR8, H1N1) and the specific HA mutants were produced in and received from the Ivanovic lab at Brandeis University, USA. For labeling of the viral membrane, an appropriate amount of virus stock was diluted to 0.5 mg/mL in PBS-EDTA (PBS with 0.2 mM NaEDTA, henceforth referred to as PBS only) in a volume of 20 uL. 0.1 μL of 0.2 mM R18 (Rhodamine C18 sodium salt, ThermoFisher) in DMSO was added, and everything was carefully mixed by pipetting up and down. After standing at room temperature for 1 h in darkness, the virus-R18 solution was added to a PBS-pre-equilibrated G-25 column (ThermoFisher) and allowed to enter the bed, 480 μL of PBS was added and the eluate discarded. Then, virus fractions per two drops were collected by eluting with 1 mL of PBS, and the fractions of most concentrated virus were combined to about 350 μL total volume. The most concentrated fractions were determined by imaging 3 μL droplets on a bare coverslip, and comprised fractions 2-6 or 3-7.

Fabs

crF8020 Fab fragments, and the AlexaFluor488-labeled version crF8020-AF488, were the same stocks as used in Otterstrom et al.,[36] stored at −80 °C. Detailed methods and characterization are described in that reference.

Coverslip and flow cell

Experiments were performed in a 5-channel PDMS-chip self-adhesing to a glass coverslip as described before. Briefly, polydimethylsiloxane (PDMS, Slygard 184; Dow Corning) was mixed with the curing agent in a 10:1 ratio, poured on a mold and allowed to harden for two days at room temperature. PDMS chips were cleaned before use by sonication with 0.5 % v/v Triton-X and 70 % ethanol for 10 min each. Cleaning involved an additional step of sonication with 1 M $NaHCO_3$ for re-use. Glass coverslips (#1.5; Marienfeld) were cleaned by sonication in detergent, acetone, ethanol (each 30 min) and 1 M KOH solution (10 min) successively, and subsequently rinsed with large amounts of water. Polyethylene

[†]This appendix was supplied by Jelle S. Blijleven from the Van Oijen Lab (Single Molecule Biophysics) at the University of Groningen.

tubing (Bioseb) provided the connection between buffer reservoirs, the flow cell and the syringe pump (NE-1000; New Era Pump Systems Inc.). Flow channel dimensions in the imaging region were 0.2 mm height × 0.5 mm width.

Proteoliposome preparation

Methods and incubation times were as described before. Liposomes were formed by extrusion through 200 nm-pore membranes (Avanti Polar Lipids), and composed of a 0.75 : 0.25 : 2.5×10^{-5} ratio of DOPC:cholesterol:biotin-DOPE (Avanti Polar Lipids). Liposomes were solubilized with Triton-X and the receptor protein GYPA (Glycophorin predominantly glycophorin A) was added in an approximate ratio of 1:27 000 GYPA:lipid, correcting for GYPA content per the manufacturer's specification sheet.

Microscope specifications

All experiments were performed on an inverted microscope, under near-total-internal-reflection conditions so that an even irradiation of the sample was assured. 488 nm and 561 nm lasers (Coherent) excited the AlexaFluor488 dyes and R18 molecules, respectively. Fluorescence was split and collected onto two halves of an EM-CCD camera (Hamatsu) to enable co-localization of Fab and viral-membrane signals.

Fusion experiment

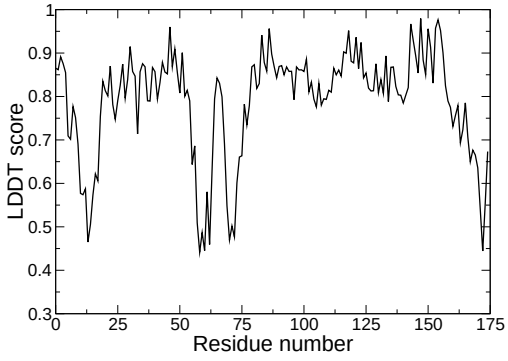
After rinsing with PBS, proteoliposomes were drawn into the channel and left for 45 min for the liposomes to spontaneously form a bilayer. The channel was then washed with PBS. Bilayer integrity was confirmed with fluorescence recovery after photobleaching (FRAP). The virus was incubated at room temperature with or without Fabs for at least 45 min to assure binding saturation. The following steps were then performed directly after eachother and with roughly the same timings: all of the virus mixture was drawn in; fluorescein-streptavidin at 0.27 μg/mL was drawn in to bind to the sparse biotin-decorated lipids in the bilayer and provide a readout of the local pH conditons; rinsing with PBS; lastly, pH 5.0 buffer (10 mM citric acid, 0.2 mM NaEDTA) was drawn in quickly to acidify the channel and trigger the viruses to fuse. Acidification typically happened at 7 to 9 min after introducing the virus into the channel and was complete within seconds. Fusion movies were taken with 200 ms frame exposure time for 4 min, and the first and last five seconds were imaged with the 488 nm laser off to allow rapid virus particle localization and fusion yield determination.

Data extraction and analysis

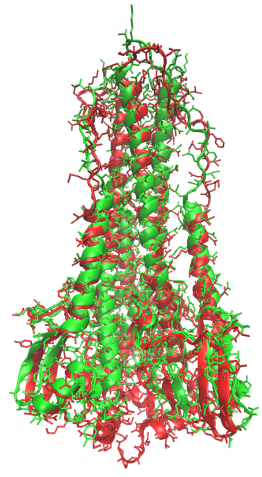
Virus particles were detected in the R18 channel with the 488 laser off by using a discoidal averaging filter, and coordinates translated to the Fab channel. The time of acidification ('pH drop') was determined by fitting an error function to the fluorescein background signal in the 488 channel. The illumination profile was corrected for by filtering out peaks and fitting a Gaussian to the whole field of view. A 7×7 pixel region of interest was used to extract R18 and Fab signal intensities per virus particle, and background was globally corrected. Hemifusion was detected from a dequenching or dissipating virus spot signal. Fusion yield, the fraction of the virus population undergoing hemifusion, was determined by counting the number of virus particles at the beginning and end of the movie without 488 illumination on, by using a threshold combined with manual inspection. Single Fab intensity was determined by flowing a low concentration ($\sim$ pM) of Fabs diluted into the same pH 5.0 buffer into a clean flow channel, rinsing, and imaging the single spots under the same conditions as in a fusion experiment. This intensity was then used to calculate the number of Fabs bound per virus particle averaged over the first 20 frames after the pH drop.

4.B Appendix: H1 homology modeling

Because there is no complete structure of H1 HA (from influenza A/Puerto Rico/8/34 H1N1) available in the database, the SWISS-Model web server (http://swissmodel.expasy.org/) was used to build the model of H1 using the H3 HA structure (from influenza A/Aichi/68-X31 H3N2, PDB entry 1HGF) as a target. The model quality, shown in Figure 4.B.1a, overall is high, except for the unstructured regions of the fusion peptide (residues 10 to 23), the B-loop (residues 57 to 75), and at the C-terminal. Only this C-terminal region is part of $HA2_{112}$ and unfavourable conformations in this region are expected to disappear during equilibration. An overlay of the resulting (unequilibrated) homology model with the (incomplete) structure of H1 is shown in Figure 4.B.1b.



(a) *Model quality.*



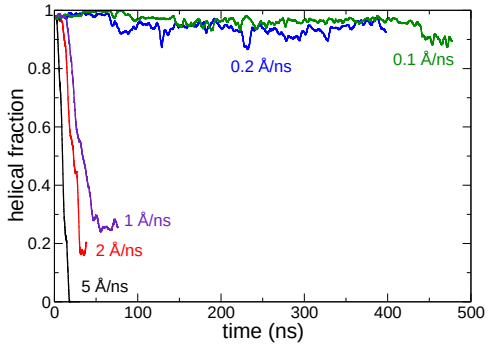
(b) *Structure alignment.*

Figure 4.B.1: *Homology modeling of H1 using the SWISS-Model web server,*[4] *with H3 (PDB entry 1HGF) as the target structure. The quality of the model is estimated by the web server using a local distance difference test (LDDT) as shown in a). In b), a cartoon of the homology model (red) is overlaid with the crystal structure of H1 (green), which only contains residues 1 to 160 of HA2 (PDB Code 1RU7).*

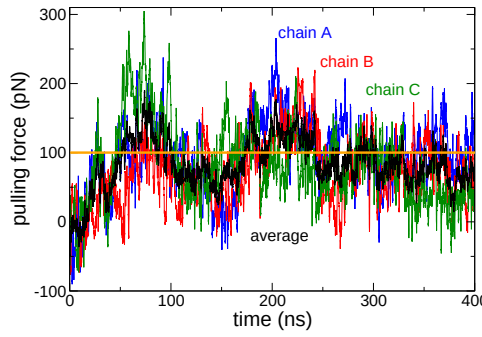
4.C Appendix: Constant pulling rate simulations

Protein unfolding can proceed either stochastically or deterministically, depending on the pulling rate.[38] At low to moderate rates, protein domains have time to spontaneously unfold under the applied load, while at high rates, these stochastic events become irrelevant and unfolding will be deterministic. Generally, the pulling force in MD simulations is in the deterministic regime to unfold the system of interest within a reasonable amount of computing time. In the present study, an 'optimal' pulling force was determined that was low enough to unfold the globular bottom without unfolding the central coiled coil. So there should be enough time for stochastic processes, such as re-establishing helical hydrogen bonds, to maintain the integrity of the coiled coil during the unfolding simulations. Additionally, stochastic breaking of hydrogen bonds is needed to let the globular bottom unfold. Exploratory pulling simulations using a model of HA that contains more of this central coiled coil, $HA2_{90}$, were used to determine this optimal pulling force.

Figure 4.C.1a shows the helical fraction of the bottom part of the coiled coil (residues 106 to 128) at different pulling rates. At rates above 1 Å/ns, the helices of the coiled coil unfold rapidly, while at a maximum pulling rate of 0.2 Å/ns, they remain more than 90% helical. The corresponding average pulling force of about 100 pN (Figure 4.C.1b) is considered to be the optimal pulling force, and was used for the pulling simulations on $HA2_{112}$ as described in the main text.



(a) *Coiled coil helix unfolding.*



(b) *Pulling force at* 0.2 Å/ns.

Figure 4.C.1: *a) Average helical fraction of the helices in the central coiled coil (residues 106 to 128) of* $HA2_{90}$ *at pH 7, while pulling at constant rates of* 5, 2, 1, 0.2 *and* 0.1 Å/ns. *b) The pulling force per chain (blue, red, green) and the average over all chains (black) for pulling rate* 0.2 Å/ns*, with an orange line drawn at the approximate average pulling force of 100 pN. All lines in a) and b) represent a* 3 ns *running average.*

4.D Appendix: Tables

Table 4.D.1: *Statistical significance of the difference between the average macroscopic unfolding times shown in Figure 4.7, as determined by a two-tailed heteroscedastic t-test.*

	H3											H1
	WT	Y162A	R163A	E128Q	E131Q	E128Q E131Q	R163A E128Q	R163A E131Q	His+	E131+	His+ E128+ E131+	WT
H3												
WT	1.000	0.860	0.001	0.839	0.001	0.000	0.003	0.008	0.397	0.003	0.000	0.000
Y162A	0.860	1.000	0.099	0.759	0.096	0.056	0.144	0.188	0.742	0.127	0.052	0.060
R163A	0.001	0.099	1.000	0.001	0.927	0.029	0.216	0.188	0.002	0.722	0.021	0.053
E128Q	0.839	0.759	0.001	1.000	0.001	0.000	0.003	0.007	0.306	0.003	0.000	0.000
E131Q	0.001	0.096	0.927	0.001	1.000	0.073	0.229	0.186	0.002	0.692	0.053	0.116
E128Q-E131Q	0.000	0.056	0.029	0.000	0.073	1.000	0.002	0.018	0.000	0.170	0.705	0.664
R163A-E128Q	0.003	0.144	0.216	0.003	0.229	0.002	1.000	0.599	0.009	0.735	0.002	0.004
R163A-E131Q	0.008	0.188	0.188	0.007	0.186	0.018	0.599	1.000	0.032	0.503	0.014	0.025
His+	0.397	0.742	0.002	0.306	0.002	0.000	0.009	0.032	1.000	0.013	0.000	0.000
E131+	0.003	0.127	0.722	0.003	0.692	0.170	0.735	0.503	0.013	1.000	0.139	0.211
His+E128+E131+	0.000	0.052	0.021	0.000	0.053	0.705	0.002	0.014	0.000	0.139	1.000	0.451
H1												
WT	0.000	0.060	0.053	0.000	0.116	0.664	0.004	0.025	0.000	0.211	0.451	1.000
S163R-N128E-K131E	0.598	0.588	0.007	0.728	0.007	0.004	0.013	0.019	0.233	0.011	0.003	0.004

Table 4.D.2: *The virus strains used for the multiple sequence alignment in Figure 4.8.*

Subtype	Accession Code	Strain
H1N1	ADA83041	A/Puerto Rico/8/1934 (PR8)
H2N2	BAG72216	A/Adachi/2/1957
H3N2	AIU46019	A/Aichi/2/1968 (X31)
H4N6	AGG81749	A/American black duck/New Brunswick/00464/2010
H5N2	ADU20361	A/American black duck/Illinois/08OS2688/2008
H6N8	AKF34349	A/American black duck/Maryland/07OS2482/2007
H7N1	AAG10656	A/African starling/England-Q/983/1979
H8N4	AEM75966	A/American black duck/Illinois/4119/2009
H9N2	BAN15811	A/African stonechat/Vietnam/8/2009
H10N6	AGG83172	A/American black duck/New Brunswick/00471/2010
H11N2	AHZ21120	A/Adelie penguin/Antarctica/178/2013
H12N6	AGG84126	A/American black duck/New Brunswick/00998/2010
H13N9	ADB46159	A/American white pelican/Minnesota/AI-07-1819/2007
H14N3	AHJ57334	A/blue-winged teal/Guatemala/CIP049H105-15/2011
H15N2	ABB90704	A/Australian shelduck/Western Australia/1756/1983
H16N3	AHM98288	A/California gull/California/1196P/2013
H17N10	AFC35438	A/little yellow-shouldered bat/Guatemala/060/2010
H18N11	AGX84934	A/flat-faced bat/Peru/033/2010